Windward environmental LLC

200 West Mercer St. • Suite 401 • Seattle, WA 98119
Phone: 206.378.1364 • Fax: 206.973.3048 • www.windwardenv.com

MEMORANDUM

To: Kris McCaig, Teck American Incorporated

From: Brian Church, Windward Environmental

Subject: Evaluation of SRC Regression Analysis of UCR Soils Data

Date: June 8, 2015

We have been asked to review SRC's May 27 memorandum entitled *Draft Summary Regression Analyses of Metal Concentrations in Residential and Upland Soils.* We have carefully reviewed the memorandum and underlying data. The memorandum's chief conclusions are that the data supports an hypothesis that aerial emissions from the Trail smelter account for lead contamination in the study area, and that the Northport smelter and local mines do not. Our preliminary analysis is that the statistical analyses are insufficient to support these conclusions and the underlying data suggest different explanatory phenomena. We comment further as follows:

1. <u>The underlying data exhibit a non-linear relationship with two marked, localized peaks in soils concentrations, which does not appear to have been adequately accounted for in the regression analyses undertaken</u>.

As a first step to a statistical analysis, it is important to plot the underlying data graphically. Otherwise, as here, the computations may represent little more than mathematical contrivances that do not reflect meaningful, linear trends in the data. That danger is perfectly illustrated here. See Attachment A (scatter plots of residential and upland soils data). Scatter plots of the residential soils data show localized peaks in metals concentrations (i.e., non-monotonic) proximate to the Boundary Transfer Station and to the immediate north of Northport in the Sheep Creek/Deep Creek area. With these two local sources removed, there is not a meaningful linear trend in the concentration data with distance from the Trail smelter; the trends are generally flat. See Attachment B.

2. <u>Although the r^2 reported are described as statistically significant, they are in fact very low, which does not support this characterization</u>.

Even if the Boundary Transfer Station and Northport-area data are included as SRC has done, the r^2 values for distance from the Trail smelter versus lead and arsenic concentrations are only 0.29 and 0.24, respectively. These are very weak correlations and are insufficient to support SRC's conclusion that there is a meaningful, linear decrease in metals in soils with distance from the Trail smelter.

SRC's regression meaning to show a relationship between soil concentrations and the historic SO_2 vegetation damage plume also drives low r^2 values. Though somewhat higher than the values derived for the distance-from-Trail analyses, the increase appears to be driven by the selection of three vegetation damage categories, which fits a regression line over a gross stair-step pattern rather than a linear or curvilinear pattern, different numbers of data points available within each category, and possibly inclusion of relict flood plain data points not reflective of aerial emissions. Also, even if demonstrable, the appearance of a statistical relationship between the SO_2 plume and soil concentrations would not be compelling evidence of the Trail smelter as a source, because the relatively localized lead and arsenic concentration patterns, likely attributable to the Boundary Transfer Station and Northport-area contamination (Exhibit B), would produce the same statistical behavior.

3. <u>The r^2 values appear to have been increased by the use of multiple independent variables, the true independence and use of which does not appear technically supported</u>.

Typically, when multiple explanatory variables are introduced in multiple linear models, a penalization is applied or understood because each variable will drive up r^2 values. That is, as variables are added to the models, an r^2 value will increase regardless of the significance or meaningfulness of each variable. As a result, one should use a penalized metric (e.g., adjusted r^2 or Akaike's Information Criterion) for evaluating the fit of a multiple linear model. The meaningfulness of any variable should be carefully considered before it is added. Here, SRC added both slope and elevation to drive an increased r^2 value for the lead model.[1] Slope and elevation would not be considered independent variables. In this geographic area, as the elevation from which a sample was taken increases, so, too, would the slope of the mountain surface from which it was taken. By including both variables in the regression analysis, SRC therefore magnifies their effect on the r^2 computation. With either or both of these variables removed, the r^2 declines, reflecting the reality depicted in the scatter plots that there appears to be no meaningful relationship between distance to the Trail smelter and the measured soil concentrations in the U.S. In SRC's other computations, slope and elevation variables appear to have been selectively included or excluded without explanation to drive r^2 results. Even then, the r^2 values are very low. They simply do not reflect a meaningful linear relationship. Also, a smelter source would be expected

[1] Specifically, the model of lead based on residential soils data.

to produce a strong curvilinear relationship between soils concentrations relative to distance from the source, which is not manifested in these data either.

4. The conclusion that soil concentrations have a negative relationship with distance from the Northport smelter is not supported by the underlying computations and relevant omitted data.

The LeRoi/Northport computation fails to consider the Weston soils concentrations dataset that measured soil concentrations prior to the Northport removal action. When that dataset is included, a trend of decreasing concentrations from the Northport smelter is in fact observed until the distance of the Boundary Transfer Station. There, the data spikes again. These spikes in metals concentrations are better explained by significant effects from two local sources, not the distant Trail smelter. By excluding the Weston dataset, SRC is merely reporting that post-removal action at Northport, the area around Boundary Transfer Station has higher concentrations generally than the area around the Northport smelter at present. Again, as the scatter plots show, that does not support the conclusion that there is a Trail impact in the study area.

5. The local mines analysis appears inapposite.

It is not clear *which* mines SRC is referring to, and not all mines would be expected to have the same magnitude of impact on soils due to their varying sizes, operational histories, containment measures, and the underlying geology which gives rise to exploration. It is likely not pure coincidence, as SRC suggests, that the several decision units from the upland soils program across the river from the Northport smelter are elevated. These decision units are not only proximate to that smelter, but they are also situated over a collection of mines, and probably also, geology that suggested mining could be viable in that area. SRC's conclusions appear to exclude without explanation the historic use, geological and geographic patterns relevant to any analyses of these data.

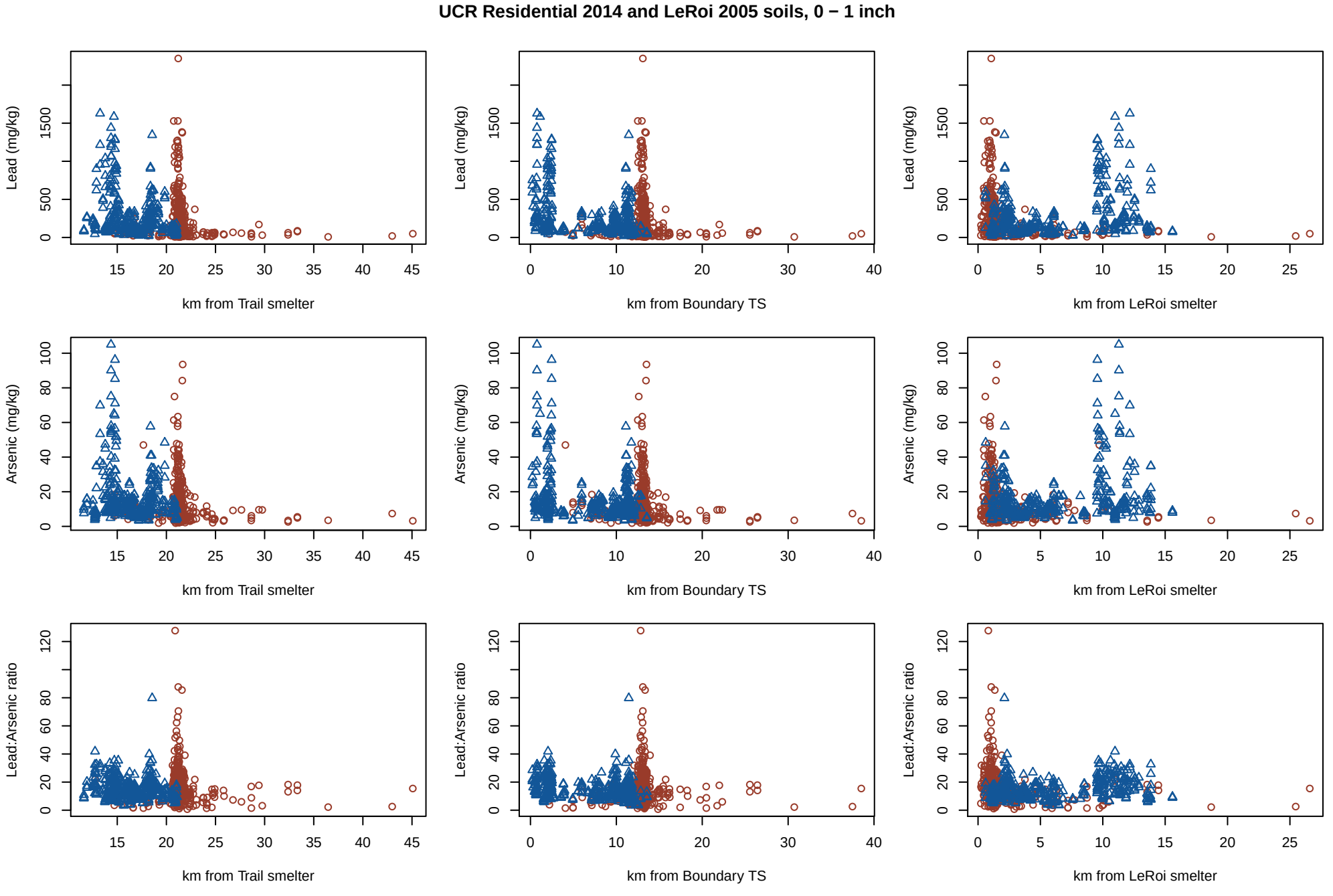
6. A multiple linear modeling should be fully explained, reproducible and analytically consistent with the underlying data.

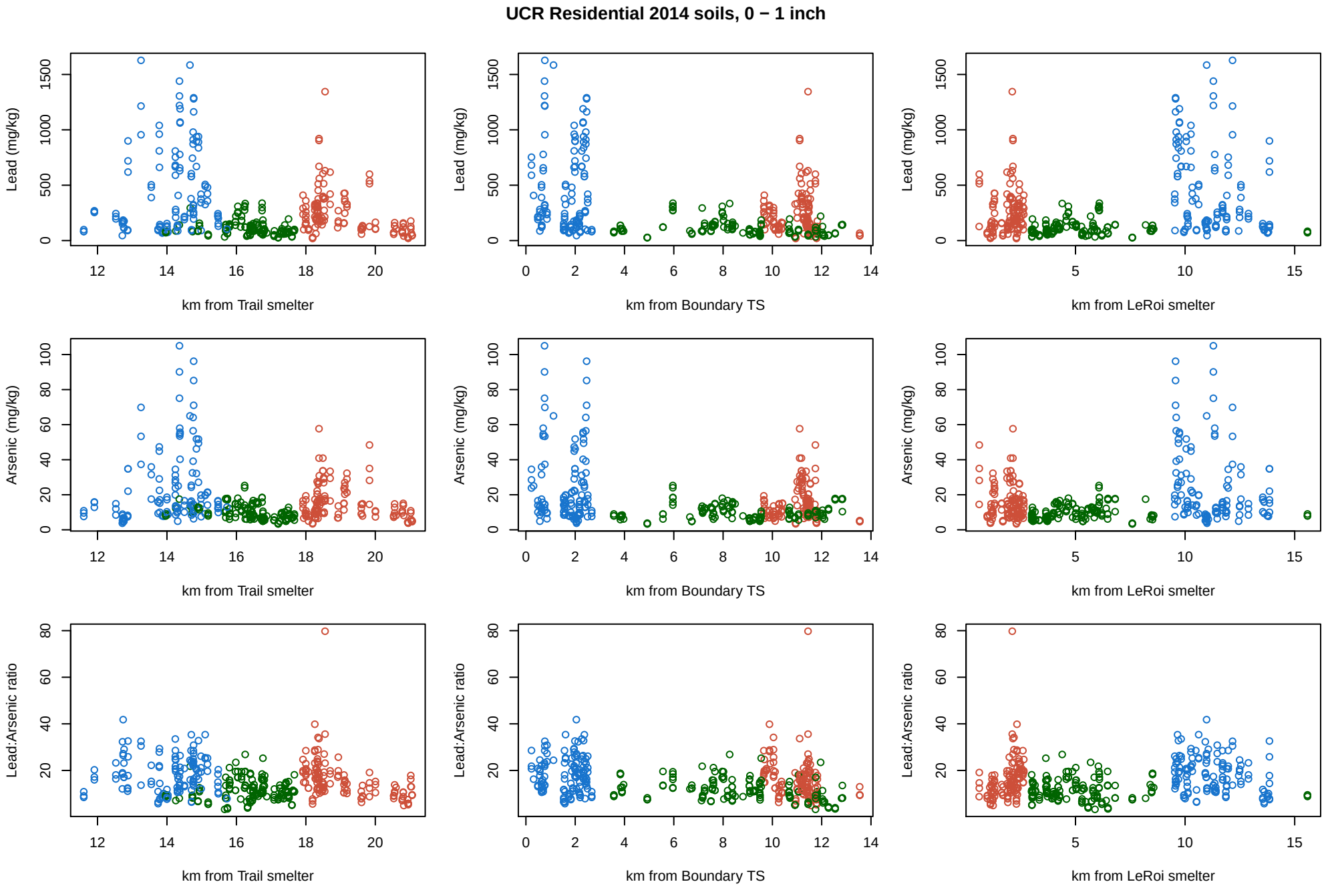
In addition to not identifying and describing the localized spikes in lead and arsenic concentrations reflected in the r^2 values resulting from their computations, SRC's multiple linear models are inconsistent with several general assumptions of analytically-sound multiple linear modeling. There is no explanation for why specific variables are included in the models or any hypotheses being tested, which casts doubt on the entire modeling process.[2] There are also no clear, linear relationships between metal concentrations and the independent variables in SRC's models. This is a violation of a primary assumption of linear modeling (i.e., linearity), undermining SRC's

[2] The assumption of relevance and relatedness between variables is not specific to linear modeling, rather it is important for interpreting statistics in general.

conclusions regarding statistical significance, the relative importance of singular variables, and reported trends in metals concentrations (e.g., change in lead or arsenic with change in river mile). Many of the variables included in the models are not normally distributed, so it is doubtful that they would collectively be multivariate-normally distributed, which is another assumption. Lastly, in reviewing the residual plots from the multiple linear models developed by SRC, there is clear heteroskedasticity of residuals, which indicates another failed assumption of each of SRC's models.

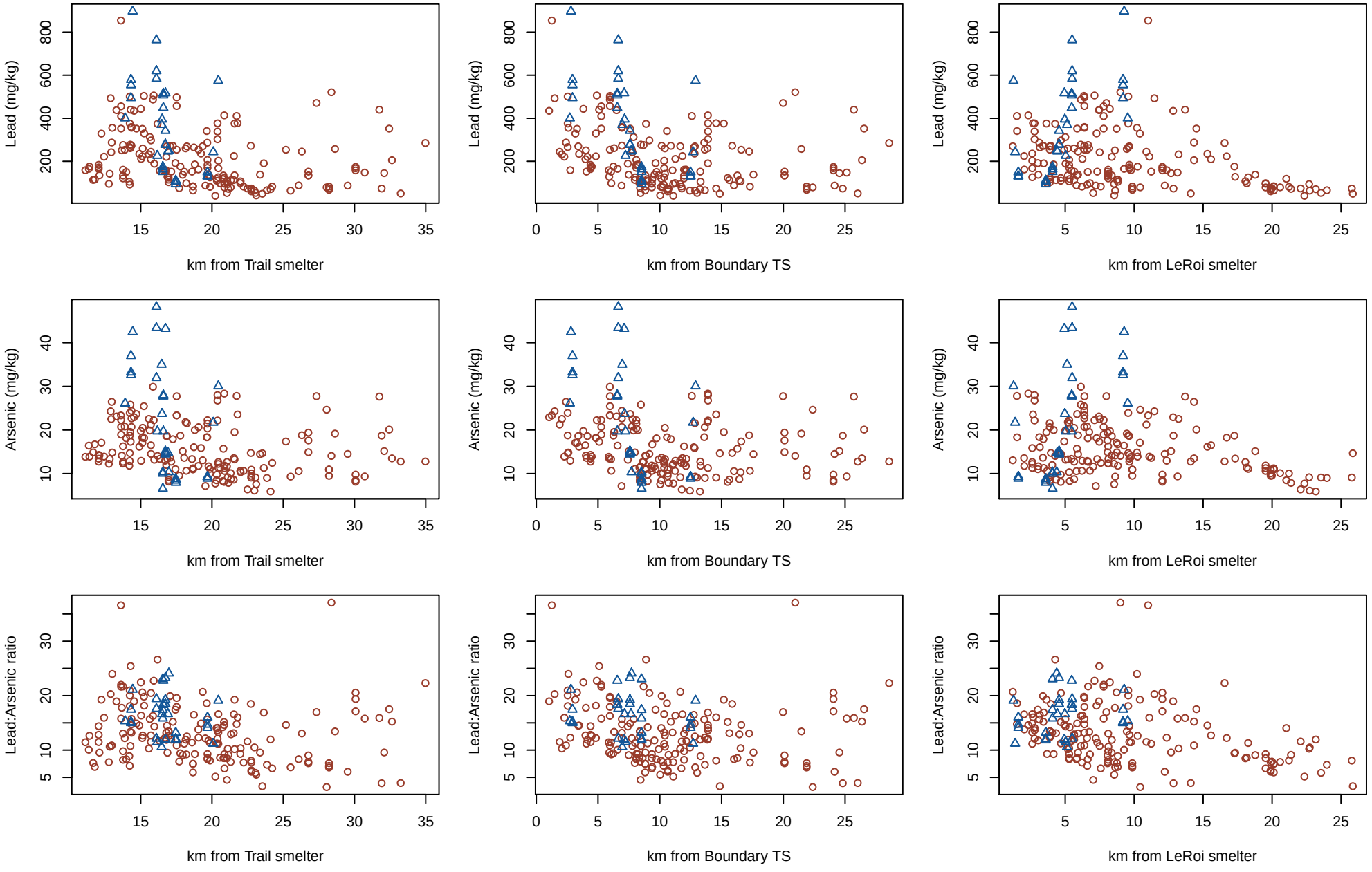
Attachment A


UCR Residential 2014 and LeRoi 2005 soils, 0 − 1 inch
Lead (mg/kg)
Arsenic (mg/kg)
Lead:Arsenic ratio
km from Trail smelter
km from Boundary TS
km from LeRoi smelter
Circles: LeRoi2005; Triangles: USEPA_2014_ResSoil



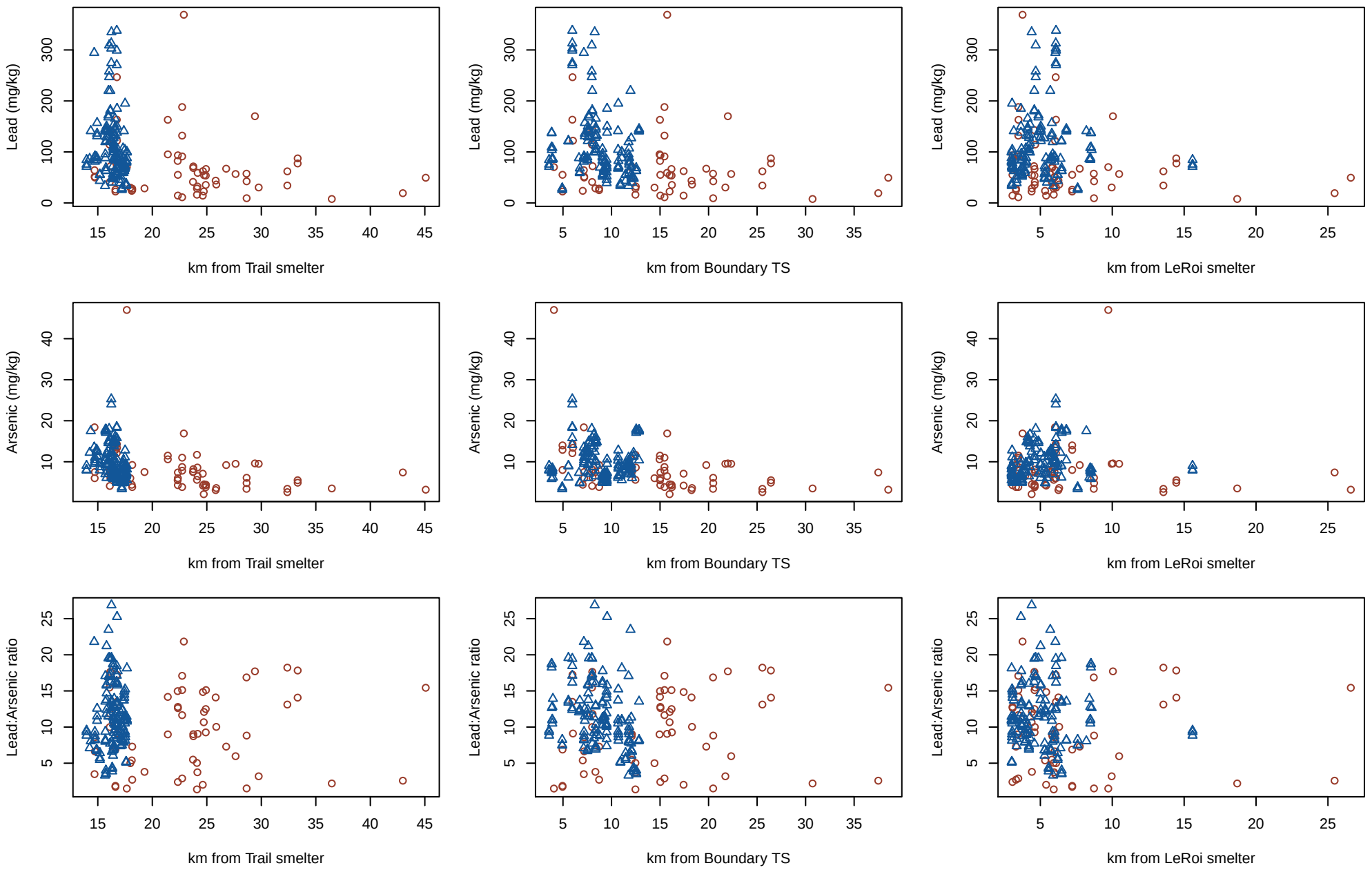
Blue: within 3 km of Boundary TS; red: within 3 km of LeRoi smelter; green: > 3 km from those two sources

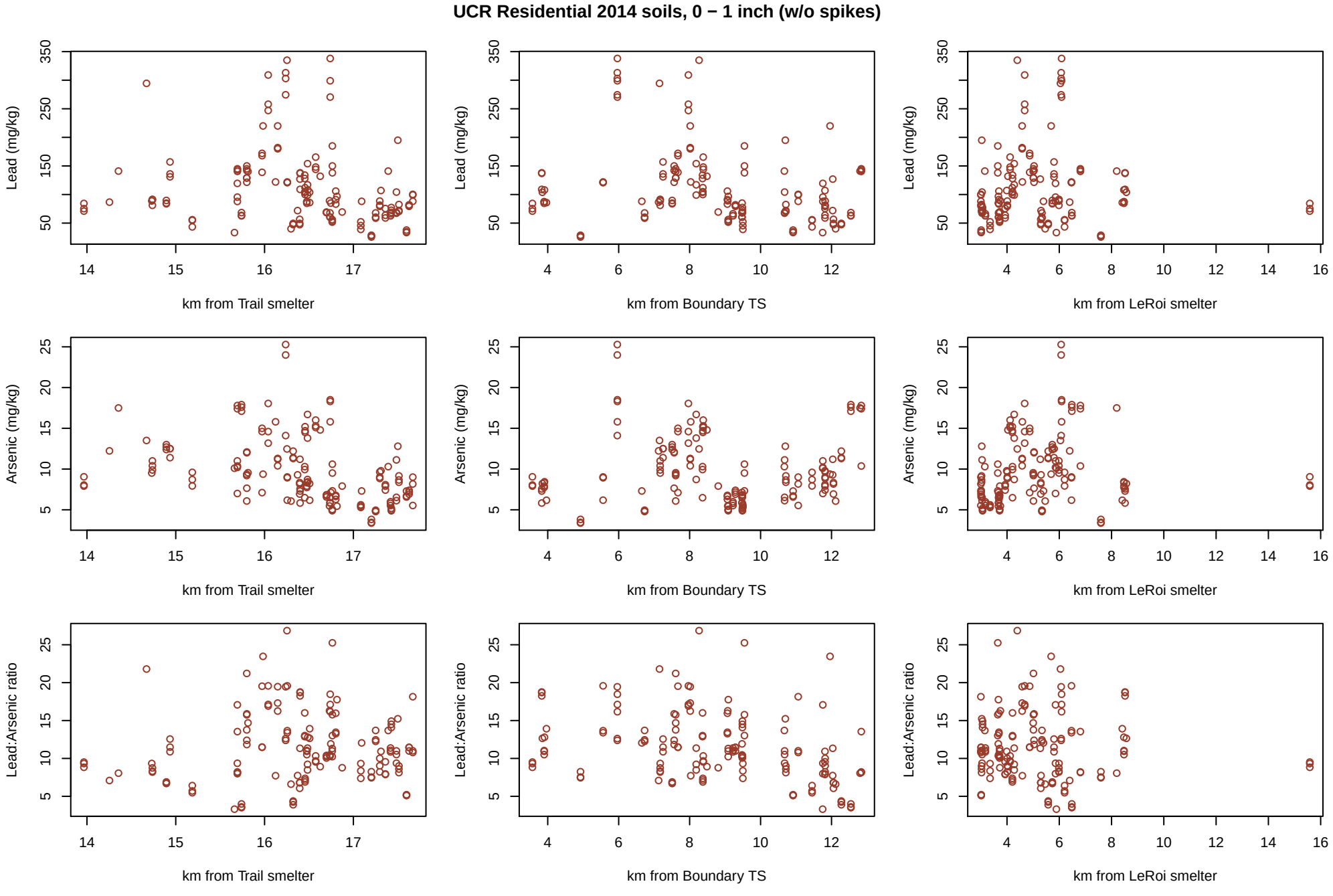
UCR Residential and UCR Upland 2014 soils, 0 − 3 inch



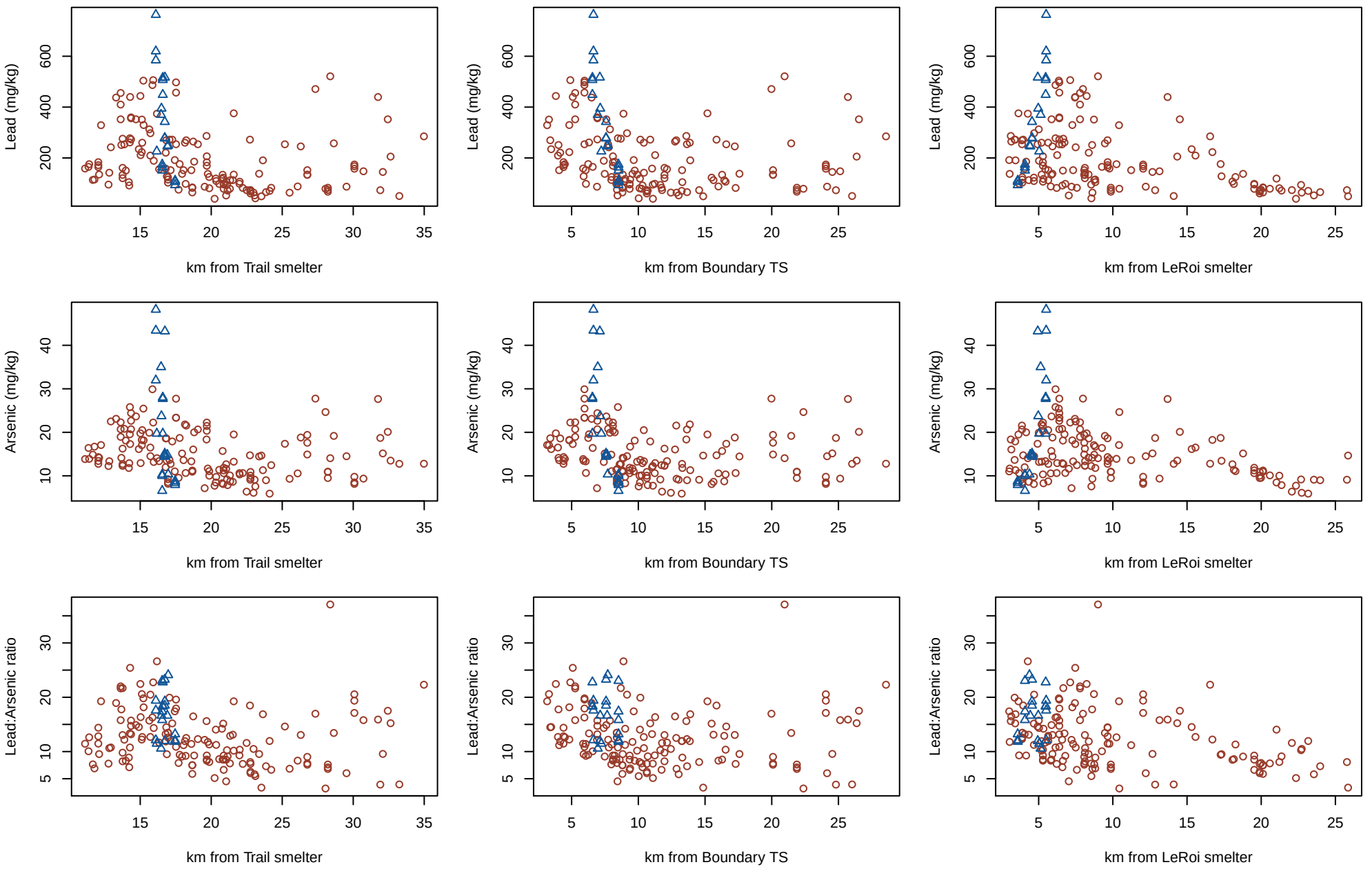
Circles: Teck_2014_UplandSoil; Triangles: USEPA_2014_ResSoil

Attachment B

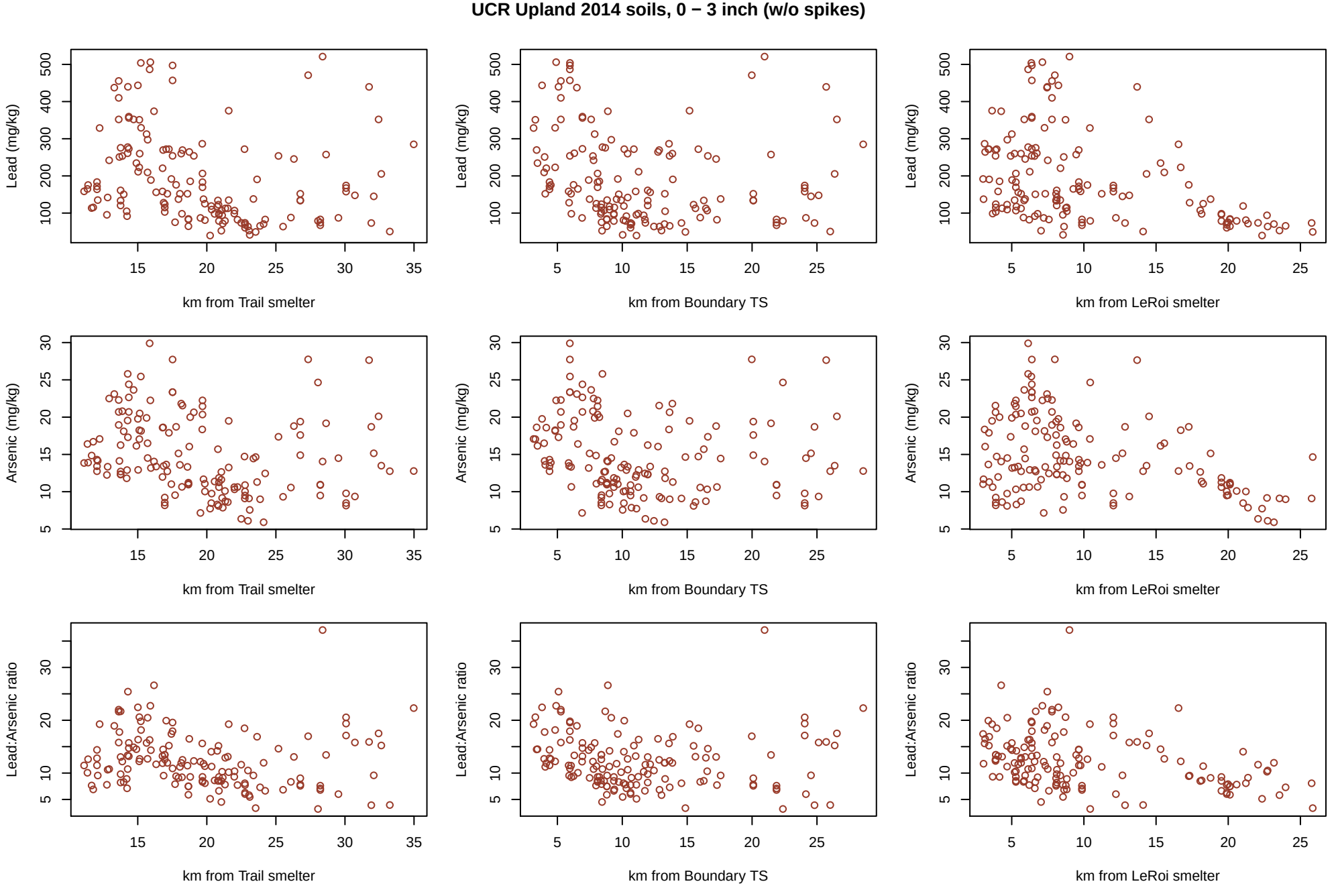

UCR Residential 2014 and LeRoi 2005 soils, 0 − 1 inch (w/o spikes)
Lead (mg/kg)
Arsenic (mg/kg)
Lead:Arsenic ratio
km from Trail smelter
km from Boundary TS
km from LeRoi smelter
Circles: LeRoi2005; Triangles: USEPA_2014_ResSoil


UCR Residential 2014 soils, 0 − 1 inch (w/o spikes)
Lead (mg/kg)
km from Trail smelter
km from Boundary TS
km from LeRoi smelter
Arsenic (mg/kg)
Lead:Arsenic ratio

UCR Residential and UCR Upland 2014 soils, 0 − 3 inch (w/o spikes)



Circles: Teck_2014_UplandSoil; Triangles: USEPA_2014_ResSoil


UCR Upland 2014 soils, 0 − 3 inch (w/o spikes)
Lead (mg/kg)
Arsenic (mg/kg)
Lead:Arsenic ratio
km from Trail smelter
km from Boundary TS
km from LeRoi smelter